Your Name: Dennis Kwadzode **Problem Set 1 Due: Feb. 18, 2025**

(1) The Poisson distribution has probability mass function

$$p(y_i|\theta) = \frac{\theta^{y_i} e^{-\theta}}{y_i!}, \qquad \theta > 0, \qquad y_i = 0, 1, \ldots \tag{1}$$

and let $y_1, \ldots, y_n$ be random sample from this distribution.

1. Show that the gamma distribution $\mathcal{G}(\alpha, \beta)$ is a conjugate prior distribution for the Poisson distribution.
2. Show that $\bar{y}$ is the MLE for θ.
3. Write the mean of the posterior distribution as a weighted average of the mean of the prior distribution and the MLE.
4. What happens to the weight on the prior mean as n becomes large?

SOLUTION

1. Conjugate Prior

The likelihood based on the sample is

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \propto \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta}.$$

Assume a Gamma prior for θ with parameters α and β:

$$p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0.$$

Multiplying the likelihood and the prior gives the posterior (ignoring normalizing constants):

$$p(\theta \mid y) \propto \theta^{\alpha-1+\sum_{i=1}^{n} y_i} e^{-(\beta+n)\theta}.$$

Since this is the kernel of a Gamma distribution, we have

$$\theta \mid y \sim \text{Gamma}\Big(\alpha + \sum_{i=1}^{n} y_i,\, \beta + n\Big).$$

Thus, the Gamma prior is conjugate to the Poisson likelihood.

2. Maximum Likelihood Estimate (MLE)

The log-likelihood is

$$\ell(\theta) = \sum_{i=1}^{n} [y_i \ln \theta - \theta - \ln(y_i!)] = \left(\sum_{i=1}^{n} y_i\right) \ln \theta - n\theta + \text{const.}$$

Differentiate with respect to θ:

$$\frac{d\ell}{d\theta} = \frac{\sum_{i=1}^{n} y_i}{\theta} - n.$$

Setting the derivative equal to zero:

$$\frac{\sum_{i=1}^{n} y_i}{\theta} - n = 0 \quad \Longrightarrow \quad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}.$$

Thus, the MLE for θ is $\bar{y}$.

3. Posterior Mean as a Weighted Average

The posterior distribution is

$$\theta \mid y \sim \text{Gamma}\Big(\alpha + \sum_{i=1}^{n} y_i,\, \beta + n\Big),$$

so its mean is

$$E[\theta \mid y] = \frac{\alpha + \sum_{i=1}^{n} y_i}{\beta + n}.$$

Since $\sum_{i=1}^{n} y_i = n\bar{y}$ and the prior mean is $\mu_0 = \alpha/\beta$, we can write:

$$E[\theta \mid y] = \frac{\alpha}{\beta + n} + \frac{n\bar{y}}{\beta + n} = \frac{\beta}{\beta + n}\left(\frac{\alpha}{\beta}\right) + \frac{n}{\beta + n}\bar{y}.$$

Thus, the posterior mean is a weighted average of the prior mean α/β and the MLE $\bar{y}$, with weights $\beta/(\beta + n)$ and $n/(\beta + n)$, respectively.

4. Behavior as n Increases

As n becomes large,

$$\frac{\beta}{\beta + n} \to 0 \quad \text{and} \quad \frac{n}{\beta + n} \to 1.$$

Thus, for large n the posterior mean is dominated by the MLE $\bar{y}$ and the influence of the prior diminishes.

(2) Consider the following two sets of data obtained after tossing a die 100 and 1000 times, respectively:

n	1	2	3	4	5	6
100	19	12	17	18	20	14
1000	190	120	170	180	200	140

Suppose you are interested in θ_1, the probability of obtaining a one spot. Assume your prior for all the probabilities is a Dirichlet distribution, where each $\alpha_i = 2$. Compute the posterior distribution for θ_1 for each of the sample sizes in the table. Plot the resulting distribution and compare the results. Comment on the effect of having a larger sample.

SOLUTION

Dirichlet Posterior for a Die Toss

We are given the following data from tossing a die:

Outcome	1	2	3	4	5	6
Counts (n=100)	19	12	17	18	20	14
Counts (n=1000)	190	120	170	180	200	140

We are interested in θ_1, the probability of obtaining a one spot. The prior for the probability vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_6)$ is a Dirichlet distribution with all parameters equal to 2:

$$\boldsymbol{\theta} \sim \text{Dirichlet}(2, 2, 2, 2, 2, 2).$$

Posterior Distribution

Because the Dirichlet prior is conjugate to the multinomial likelihood, the posterior distribution is

$$\boldsymbol{\theta} \mid y \sim \text{Dirichlet}(2 + n_1,\, 2 + n_2,\, \ldots,\, 2 + n_6).$$

Marginal Posterior for θ_1

The marginal posterior for θ_1 is a Beta distribution:

$$\theta_1 \mid y \sim \text{Beta}\Big(2 + n_1,\, \sum_{i \neq 1}(2 + n_i)\Big).$$

For $n = 100$:

$$n_1 = 19 \quad \Longrightarrow \quad 2 + n_1 = 21.$$

For the other outcomes:

$$2 + 12 = 14, \quad 2 + 17 = 19, \quad 2 + 18 = 20, \quad 2 + 20 = 22, \quad 2 + 14 = 16.$$

The second parameter is:

$$14 + 19 + 20 + 22 + 16 = 91.$$

Thus, for $n = 100$:

$$\theta_1 \mid y \sim \text{Beta}(21, 91).$$

For $n = 1000$:

$$n_1 = 190 \implies 2 + n_1 = 192.$$

For the other outcomes:

$$2 + 120 = 122, \quad 2 + 170 = 172, \quad 2 + 180 = 182, \quad 2 + 200 = 202, \quad 2 + 140 = 142.$$

The sum for the remaining outcomes is:

$$122 + 172 + 182 + 202 + 142 = 820.$$

Thus, for $n = 1000$:

$$\theta_1 \mid y \sim \text{Beta}(192, 820).$$

Plotting the Posterior Densities

Below is a plot comparing the posterior densities of θ_1 for the two sample sizes. The blue curve corresponds to the $n = 100$ case (Beta(21,91)) and the red curve to the $n = 1000$ case (Beta(192,820)).

[domain=0,0.4, samples=100, xlabel=θ_1, ylabel=Density, legend pos=north east, title=Posterior Density for θ_1] [blue, thick] (x$^(21-1)*(1-x)^(91-1))/(exp(lngamma(21)+lngamma(91)-lngamma(112)));$Beta$(21,91);$ [red, thick] (x$^(192-1)*(1-x)^(820-1))/(exp(lngamma(192)+lngamma(820)-lngamma(1012)));$Beta$(192,820);$

Figure 1: Posterior densities for θ_1 for $n = 100$ (blue) and $n = 1000$ (red).

Discussion

Both posteriors are centered near the observed proportion of ones:

$$\hat{\theta}_1 \approx \frac{19}{100} = 0.19 \quad \text{and} \quad \frac{190}{1000} = 0.19.$$

However, the posterior for $n = 1000$ is much more concentrated around 0.19, indicating a smaller variance due to the larger sample size. This illustrates that as the sample size increases, the data increasingly dominate the prior, resulting in a more precise (less variable) estimate of θ_1.